

Une plateforme de management des connaissances pour le domaine des ressources spatiales

Cédric Pruski*, Louis Deladiennée*, Emmanuel Scolan*, Marcos Da Silveira*

* Luxembourg Institute of Science and Technology
5, avenue des Hauts-Fourneaux, L-4362, Esch-sur-Alzette, Luxembourg
{cedric.pruski, louis.deladiennee, emmanuel.scolan,marcos.dasilveira}@list.lu

Résumé. L'intérêt toujours croissant aussi bien du monde académique que de l'industrie et des institutions gouvernementales pour les informations sur les ressources spatiales a mis en évidence la difficulté de trouver, d'accéder à, d'intégrer, et de réutiliser ces informations. Toute cette connaissance est diffusée dans des publications scientifiques, des brevets, des news, et des rapports, qui sont régulièrement mis à disposition de la communauté sur le Web à travers de nombreux médias et dans différents formats. En collaboration avec l'agence spatiale européenne (ESA), nous développons actuellement une plateforme de management des connaissances du domaine des ressources spatiales. Cet outil, reposant sur l'utilisation des technologies du Web Sémantique telles que les graphes de connaissance et les ontologies, a pour vocation de formaliser la connaissance du domaine contenue dans les documents évoqués précédemment et de les mettre à disposition de la communauté. Dans cet article applicatif, nous décrivons les concepts et techniques pour l'extraction et la gestion des connaissances mis en œuvre pour le design et l'implémentation d'une telle plateforme.

1 Introduction

La colonisation durable de l'espace nécessite l'utilisation et l'exploitation de ressources spatiales telles que les minéraux, l'eau, et les matériaux volatiles (Rovetto, 2017; Meurisse et al., 2021; ISECG, 2021). De nombreuses agences spatiales visent la production et l'utilisation in situ de ces ressources pour notamment organiser l'exploration humaine à long terme de la Lune, de Mars, et de l'espace¹ (ESA, 2019). Cette quête est largement soutenue par un certain nombre d'initiatives nationales et internationales (cf. le programme Artemis), par le développement de nouvelles technologies, par l'élaboration du cadre politique et juridique, par le financement de projets de recherche scientifique, ainsi que la mise en place d'activités commerciales. Les États-Unis et le Luxembourg ont promulgué des lois sur l'exploration pacifique et l'utilisation durable des ressources spatiales au profit de l'humanité, faisant d'eux des leaders mondiaux dans ce domaine². En collaboration avec l'Agence spatiale européenne (ESA), l'initiative luxembourgeoise a conduit en 2020 à la création d'un centre d'innovation dédié

1. <https://www.globalspaceexploration.org>

2. <https://space-agency.public.lu/en/space-resources.html>

aux ressources spatiales, l'European Space Resources Innovation Centre (ESRIC)³. L'ESRIC a pour ambition de devenir le centre d'expertise internationalement reconnu pour les aspects scientifiques, techniques, commerciaux, et économiques liés à l'utilisation des ressources spatiales pour l'exploration humaine et robotique, pour le développement d'une économie pour le domaine spatial. Dans cet objectif, l'ESRIC s'efforce de connecter et de fédérer la communauté des ressources spatiales en créant un environnement ouvert et collaboratif pour encourager le dialogue, l'échange d'idées, et les coopérations scientifiques, technologiques, et économiques.

Dans ce contexte, nous développons actuellement une plateforme pour l'extraction, la gestion, et le partage des connaissances pour les ressources spatiales⁴ (Da Silveira et al., 2022). Il s'agit d'une application web qui propose un ensemble de services pour interroger, analyser, et visualiser les connaissances du domaine des ressources spatiales obtenues en extrayant automatiquement le contenu d'articles scientifiques, de brevets, de news et plus généralement de documents du Web pertinents pour le domaine des ressources spatiales. Cette plateforme vise à fédérer la communauté des ressources spatiales en offrant la possibilité d'accéder et de contribuer à l'état de l'art des connaissances scientifiques, techniques, commerciales, et juridiques dans le domaine. Cet outil permettra également à ses utilisateurs d'obtenir des informations actualisées sur les événements régulièrement publiés sur les réseaux sociaux. La communauté du domaine des ressources spatiales ne cesse de croître et par conséquent, la demande pour cet outil augmente également. Cette communauté est actuellement composée d'entreprises, d'universités, d'organismes et d'agences de recherche et de technologie, ainsi que d'organisations non industrielles publiques et privées. Beaucoup de ces acteurs clés (près de 40% selon une enquête menée par l'ESRIC en 2021) ne sont pas encore actifs dans le secteur spatial, mais recherchent des opportunités dans ce domaine, soulignant l'impact potentiel de ce type de plateforme. Dans cet article applicatif, nous décrivons l'utilisation des concepts et techniques de l'extraction et du management des connaissances pour le design et l'implémentation d'une plateforme de management des connaissances pour le domaine des ressources spatiales développée pour répondre aux besoins de la communauté. Cet outil repose sur un graphe de connaissance pour la représentation des connaissances du domaine construit sur la base d'une ontologie du domaine conçue suivant les bonnes pratiques du Web Sémantique et d'un ensemble de services accessibles à partir du Web pour le partage et l'exploitation de ces connaissances.

L'article est structuré comme suit : La section 2 présente les travaux existants similaires aux nôtres. La section 3 porte sur les motivations de ces travaux. La section 4 décrit l'architecture, les principaux composants et services offerts par la plateforme. Ceci inclus une ontologie du domaine (section 4.1) et le graphe formalisant les connaissances du domaine (section 4.2). La section 5 décrit l'implémentation de la plateforme. La section 6 décrit l'évaluation des composants de la plateforme. Enfin, la section 7 sert de conclusion et donne une idée des travaux futurs.

2 Travaux relatifs

Dans cette section nous décrivons les applications similaires à la nôtre pour mettre en évidence le besoin d'une telle plateforme, ainsi que le bien fondé des concepts et de la technologie

3. www.esric.lu

4. <https://knowledge.esric.lu>

mis en œuvre. Une première famille d’approches comparables se focalise comme nous sur le domaine spatial, puis d’autres se focalisent sur d’autres domaines ou sur certains types de documents où de sources d’information.

L’étude SLOAN est une initiative connue pour le domaine (York et al., 2000). A travers cette application, ont été créées les cartes tridimensionnelles les plus détaillées de l’Univers. Ceci inclut des images multicolores profondes d’un tiers du ciel et des spectres pour plus de trois millions d’objets astronomiques. Les données de la plateforme SDSS sont accessibles au public dans un format utile à un large éventail d’utilisateurs allant des étudiants débutants aux astronomes amateurs et professionnels. La Space platform (<https://spacepp.com>) offre des informations sur certains composants de l’écosystème du domaine spatial. Les données sont fermées et portent surtout sur les entités géant des satellites et des lanceurs. Cette plateforme vise principalement un public composé de startups, de PME, de grandes entreprises, d’investisseurs, d’universitaires, et d’agences spatiales et offrent des services payant pour identifier des appels à projet, des levées de fonds et des appels d’offre ouverts. The Space Resource (<https://www.thespaceresource.com>) a été créée en 2018 en tant que plateforme médiatique indépendante dédiée à la création d’une communauté interactive de passionnés et de professionnels des ressources spatiales. En termes d’information, la couverture se concentre sur le partage d’idées et de nouvelles, visant à accroître la collaboration entre les nombreux professionnels travaillant à révolutionner le développement des ressources spatiales. Le site Web a été mis à jour pour la dernière fois en septembre 2021. Notons que le contenu n’est pas exprimé dans un format interprétable par l’ordinateur. Enfin, il existe un certain nombre de sites Web ou de base de données ciblant un aspect très précis des ressources spatiales. On retrouve parmi ces outils la CSM Planetary Simulant Database (<https://simulantdatab.com/>) fournissant des informations sur les simulants, Factories in Space (<https://www.factoriesinspace.com>) pour l’économie spatiale émergente et la fabrication dans l’espace et enfin la famille de Wiki tel que lunarpedia, spacepedia et marspedia dans l’idée de Wikipedias dédiés aux connaissances sur des astres précis. Les initiatives existantes ne couvrent donc que partiellement le domaine spécifique des ressources spatiales.

L’approche que nous proposons s’appuie sur les concepts et les technologies du Web Sémantique telles que les graphes de connaissances et les ontologies pour intégrer, formaliser, partager et générer les connaissances du domaine des ressources spatiales. Dans le paragraphe ci-après, nous analysons les approches similaires, pas nécessairement orientées vers le spatial, et leurs originalités.

SpaceML (Koul et al., 2020) regroupe une suite d’outils développée par un consortium international ayant pour but de permettre d’utiliser le résultat de projets de recherche en Intelligence Artificielle pour le domaine spatiale. Parmi ces résultats, on retrouve un projet visant à utiliser les graphes de connaissances pour formaliser la connaissance contenue dans des articles scientifiques portant sur la gestion de l’hydrogène⁵. Les auteurs combinent également des approches d’IA sub-symbolique avec ce graphe de connaissances afin de proposer une classification par ordre d’intérêt pour les utilisateurs. Ces travaux ne décrivent pas les sources d’informations utilisées pour construire ce graphe et se restreignent à un domaine très précis. Dans les travaux de Auer et al. (2021), les auteurs décrivent la construction d’un graphe volumineux des connaissances du monde académique structuré par discipline obtenues via CrossRef (Hendricks et al., 2020). De par la grande quantité de données traitée, ORKG s’appuie sur des

5. <https://vimeo.com/751598254/b05e37e643>

méthodes de crowdsourcing pour assurer une bonne qualité des données et offre également un ensemble de services pour l'exploitation du graphe. Nos deux approches sont complémentaires dans la mesure où nous traitons d'un sous-domaine du monde spatial pouvant étendre ORKG alors que ORKG vient compléter notre graphe par des connaissances de domaines connexes au notre (e.g. chimie, physique, géologie). Plus récemment, les travaux de Liu et al. (2022) visent également la création d'un graphe pour la représentation des connaissances du monde académique avec comme unique point d'entrée le portail IOSpress⁶. Le fait d'avoir une seule source d'information permet d'éviter les doublons provenant de plusieurs sources différentes comme c'est le cas pour notre plateforme.

Cette revue de la littérature montre d'une part qu'une telle plateforme de partage des connaissances n'existe pas spécifiquement pour le domaine des ressources spatiales et d'autre part que les outils comparables se basent sur un nombre restreint de sources de même nature (articles scientifiques) et, par conséquent, n'offrent qu'une vision académique du domaine. La cartographie d'un domaine appliqué comme celui des ressources spatiales nécessite la collecte élargie des connaissances aussi bien académiques que technologiques, légales, économiques, politiques, éducatives, et même événementielles en ciblant une plus grande variété de sources d'information pour construire un graphe pertinent du domaine.

3 Motivations

Le management des connaissances est une activité primordiale pour le développement durable d'activités le long de la chaîne de l'innovation (science-technologie-business/marché), mais aussi dans les domaines de l'économie, de la législation, et de la politique. En plus d'éviter la perte de savoir, la gestion des connaissances constitue un vecteur fédérateur de communautés de pratique.

Le domaine des ressources spatiales est encore un secteur en voie de définition et dont les limites peuvent fluctuer selon les visions et intérêts. Une enquête menée en 2021 par l'ESRIC auprès de sa « communauté des ressources spatiales » a cependant permis d'en dessiner quelques contours. Cette communauté se situe tout au long de la chaîne de l'innovation (académie - centres d'innovation - industrie) et comprend aussi bien des organisations publiques, gouvernementales, que privées. 40% ne sont pas encore impliqués dans une activité spatiale. L'intérêt de cette communauté s'étend à l'ensemble de la chaîne de valeur des ressources spatiales (ISECG, 2021). La communauté des ressources spatiales est donc très variée et multiple dans ses intérêts et attentes. Néanmoins, l'attente principale concerne l'accès et le partage de la connaissance académique et technologique du domaine.

Suivant le résultat de cette enquête, il est apparu qu'un des outils les plus adéquats pour atteindre les objectifs du management des connaissances de l'ESRIC est un portail de connaissances sur le secteur des ressources spatiales. Un tel portail devra être basé sur :

- une ontologie cartographiant les concepts du domaine,
- l'accès à des informations/connaissances appropriées et à jour (i.e. base documentaire requêtable, agenda d'événements, newsletters, alertes),
- des fonctionnalités supportant des communautés de pratique permettant ainsi l'échange d'informations (i.e. blogs, événements),

6. <https://www.iospress.com>

- la possibilité de contribuer au partage de ces informations/connaissances, et de collaborer (i.e. Wikipedia pour les ressources spatiales).

L'implémentation incrémentale d'un tel portail inclut i) la création et la mise à jour régulière d'une ontologie, ii) la connexion à plusieurs sources de données, d'informations et de documents dont le contenu est capturé sous forme de graphes de connaissance, iii) le développement d'outils d'analyses de données adaptés pour traiter de grandes quantités d'informations et leurs évolution, iv) la mise en place des fonctionnalités attendues par la communauté, v) tout en garantissant la qualité et la sécurité des accès aux informations et connaissances partagées.

4 Architecture de la plateforme

Les besoins de la communauté décrits à la section précédente montrent la nécessité pour notre plateforme d'être en capacité :

- d'extraire l'information pertinente de sources d'information fiables,
- de formaliser cette information et de faire émerger sa sémantique afin de la transformer en connaissance,
- d'exploiter cette connaissance à travers un ensemble de services adéquats

Pour répondre à ces besoins, nous nous appuyons sur les technologies du Web Sémantique qui offrent des moyens standardisés pour décrire la sémantique des concepts d'un domaine grâce à des ontologies (Gruber, 1993), et pour lier ces concepts à la connaissances capturée dans un graphe de connaissances (Bizer et al., 2011). Dans cette section nous décrivons la construction et l'utilisation d'une ontologie du domaine et d'un graphe de connaissance pour répondre aux besoins de la communauté des ressources spatiales (voir FIG. 1).

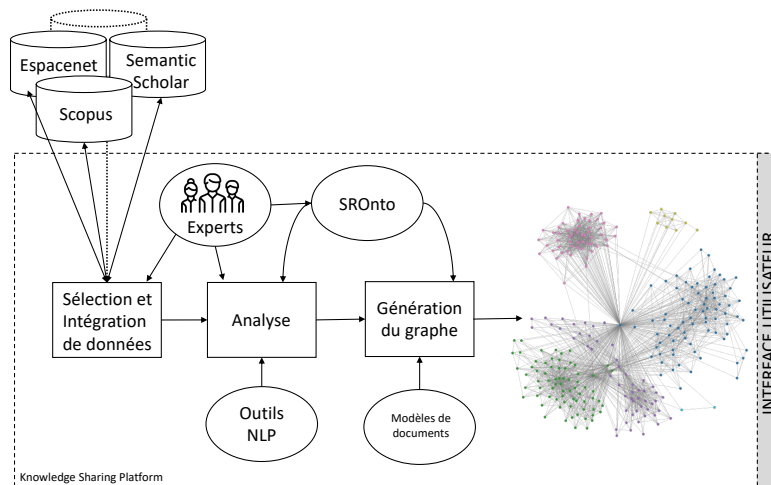


FIG. 1 – Composants et activités de la plateforme KSP.

4.1 SROnto : une ontologie du domaine des ressources spatiales

La construction de la plateforme KSP et notamment les processus d'intégration et de partage des connaissances s'appuie essentiellement sur une ontologie du domaine, SROnto, que nous avons construite en étroite collaboration avec des experts du monde des ressources spatiales et d'autres du domaine de l'ingénierie des connaissances. Cette ontologie est construite suivant les principes FAIR (Wilkinson et al., 2016) et les bonnes pratiques du Web Sémantique tel que la méthode NeON (Suárez-Figueroa et al., 2012) avec l'ambition de devenir un standard du domaine. En effet, seuls des glossaires comme JORC (Stephenson, 2001), contenant certains termes du domaine existent actuellement. Cependant, ceux-ci ne sont pas formalisés de façon compréhensible par des machines, prévenant ainsi l'utilisation de logiciels intelligents. SROnto sert non seulement à décrire les connaissances du domaine, mais est également utilisée pour, d'une part, annoter les documents ingérés dans notre graphe de connaissance (voir la section 4.2) et, d'autre part, enrichir les requêtes émises par les utilisateurs au moment d'interroger la plateforme (Pruski et al., 2011) sur la base des relations sémantiques existants entre les termes de la requête et les labels de concept de l'ontologie.

Une ontologie repose principalement sur deux aspects : la partie terminologique et la partie logique. Pour SROnto, la terminologie a été déterminé d'une part en utilisant les glossaires existants (e.g., LORS 101, JORC) fournissant les termes et leurs définitions et, d'autre part, par l'extraction manuelle par les experts du domaine des termes d'un ensemble de rapports rédigés par les acteurs clés du secteur des ressources spatiales. L'ensemble des termes ainsi identifiés va constituer les labels des concepts de SROnto. Pour la partie logique (i.e., l'axiomatisation de l'ontologie) et notamment la hiérarchisation des concepts, nous nous sommes appuyés sur la réutilisation d'un ensemble d'ontologies existantes. La partie supérieure de SROnto (i.e., la partie la plus abstraite) repose entièrement sur Basic Formal Ontology (Arp et al., 2015) pour faciliter la compréhension de nos concepts qui étendent ceux de BFO et leur future réutilisation. La partie intermédiaire de SROnto repose sur l'importation de concepts définis dans d'autres ontologies existantes telles que ENVO pour décrire des connaissances environnementales (e.g., le type de sol où trouver de la ressource), CHEBI pour la description d'éléments chimiques (comme l'oxygène contenu dans la régolithe lunaire) ou OBI pour la description des processus tels que l'extraction de l'oxygène de la roche. La partie la plus spécifique de SROnto contient les concepts spécifiques au domaine des ressources spatiales que nous avons ajoutés sur la base des termes extraits précédemment. Lors de la formalisation de ces concepts, nous avons enrichi leur description par des métadonnées standards provenant par exemple du Dublin Core afin de renforcer les aspects FAIR de notre ontologie. SROnto est exprimé suivant le standard OWL et contient finalement 556 classes mais aussi 159 Object et Data properties et 59 instances de concepts. De plus, notre ontologie est accessible au public à partir de l'URL permanente <https://purl.org/esric/sronto>. L'aspect *Findable* des principes FAIR été renforcé via l'enregistrement de SROnto dans le référentiel Linked Open Vocabulary (Vandenbussche et al., 2017). Nous avons associé la ressource à une licence creative commons pour faciliter la réutilisation de l'ontologie. De plus, la documentation de l'ontologie a été générée à l'aide de l'application logicielle WIDOCO (Garijo, 2017), ce qui la rend compréhensible par l'homme.

SROnto étant au cœur de notre plateforme KSP, il est par conséquent crucial pour la communauté spatiales d'y intégrer en permanence de nouvelles connaissances dans le graphe et de la faire évoluer en conséquence. C'est pourquoi nous avons mis en place un processus de

maintenance régulière de notre ontologie. Ce dernier commence par une analyse automatique du contenu du graphe de connaissance de la KSP en utilisant un ensemble d'outils de traitement automatique des langues capables d'extraire de nouveaux termes et de nouvelles relations non présents dans la version actuelle de SROnto. Le résultat de cette analyse automatique est ensuite revu et enrichi par un comité composé d'experts du domaine des ressources spatiales et des gestionnaires de la plateforme. Ils évalueront les nouveaux termes et suggéreront des améliorations en fonction de leur connaissance personnelle du domaine (connaissance tacite). Le résultat est ensuite transmis aux ingénieurs de la connaissance qui formaliseront le contenu fourni selon les bonnes pratiques et standards du W3C et générera une nouvelle version de SROnto, qui est ensuite validée par le comité suivant un processus itératif jusqu'à l'obtention d'un consensus. La version validée de l'ontologie est publiée sur le Web, ajoutée dans un graphe de connaissance historique (Cardoso et al., 2020) et utilisée par la KSP afin d'ingérer du nouveau contenu. Le pipeline de production garantit qu'une fois la nouvelle version de SROnto publiée, la documentation de l'ontologie est également mise à jour.

4.2 Graphe de connaissances du domaine des ressources spatiales

Le graphe de connaissance est la deuxième composante de notre plateforme KSP. Comme son nom l'indique, il permet de représenter sous forme de graphe un ensemble de métadonnées associées aux entités pertinentes pour les besoins de la communauté (i.e., des documents, les acteurs clés du domaine, les événements importants). Les besoins de la communauté, évoqués à la section 3, nous ont poussés à exprimer dans notre graphe les notions de :

- Personnes (*Person* sur la FIG. 2) afin de représenter les auteurs, éditeurs, inventeurs (pour les brevets) ou autre entité humaine jouant un rôle par rapport aux documents traités par la plateforme.
- Organisation (*Organisation* sur la FIG. 2) pour représenter les instituts de recherche, les universités, les entreprises ou encore les organisations gouvernementales ou non-gouvernementales (ou autres) du domaine des ressources spatiales. Ces organisations sont reliées aux personnes par la relation «AFFILIATED_TO».
- Documents (*Documents* sur la FIG. 2) pour décrire les articles scientifiques, les brevets, les rapports techniques ou les documents légaux. Cette entité est centrale dans notre graphe car elle relie les personnes, les organisations, le domaine et les événements à travers 5 relations. Cette notion est davantage raffinée car un ensemble de propriétés de ces documents sont spécifiés (e.g., le type de document, le DOI, etc).
- Domaine (*Domain* sur la FIG. 2) permet de spécifier à quel domaine un document appartient ou à quel domaine une personne contribue. La notion de domaine permet de faire le lien entre les informations du graphe et notre ontologie car un concept de SROnto est vu comme un domaine dans notre cas. Ce lien entre les concepts de l'ontologie et les documents sont établis grâce à des outils du traitement de la langue naturelle sur la base du résumé des documents et de leur titre.
- Événement (*Venue* sur la FIG. 2) fait référence aux événements comme les conférences, symposium ou colloques pertinents pour le domaine des ressources spatiales.

Au moment de la rédaction de cet article, le contenu de notre graphe de connaissance et obtenu par l'interrogation via des APIs dédiés, l'analyse et la restructuration des informations provenant :

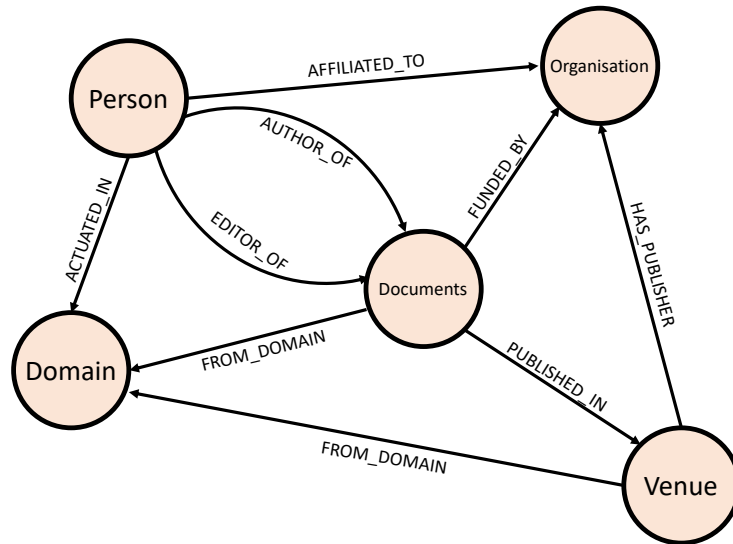


FIG. 2 – Métamodèle du graphe de connaissance.

- du corpus du portail Semantic Scholar (Lo et al., 2019) offrant accès à de la littérature scientifique.
- de la base de connaissance de Scopus (Burnham, 2006) donnant accès aux documents scientifiques dans les domaines de la science, de la technologie, de la médecine, des sciences sociales, des arts et des sciences humaines.
- de la base de brevet fournie par le portail Espacenet⁷ afin de couvrir la partie technologique important pour la communauté.

Dans un futur proche, nous comptons également ajouter d'autres sources d'information comme des sites de news pour enrichir notre graphe.

Dans notre contexte, une difficulté majeure dans l'intégration d'information provenant de plusieurs sources différentes consiste en l'identification d'information redondante. Ceci survient notamment lorsque deux bases de connaissances distinctes nous fournissent des informations sur le même document ou les mêmes personnes par exemple. Nous avons donc mis en place un mécanisme simple qui consiste à comparer, pour deux documents, si ces derniers ont le même titre ou le même identifiant (DOI) afin d'en déduire qu'il s'agit bien du même document. Une fois identifié, les informations provenant des deux sources sont ajoutées au graphe afin d'avoir une connaissance la plus complète possible.

5 Implémentation de la plateforme

La plateforme KSP est une application Web (<http://knowledge.esric.lu>) offrant actuellement les fonctionnalités suivantes :

7. <https://worldwide.espacenet.com>

- Une interface afin d’interroger la base de connaissance. Ceci peut se faire directement en cliquant sur des mots clés proposés provenant de SROnto soit à travers de l’interface de saisie et l’utilisation de mots clés entrés par l’utilisateur.
- Un accès direct à l’ontologie SROnto et à la définition de ses concepts.
- Un lien vers la liste des fournisseurs de simulants de régolithe extraite des rapports (ISECG, 2021).

D’un point de vue implémentation, les composants dépeints sur la FIG. 3 forment la plateforme KSP.

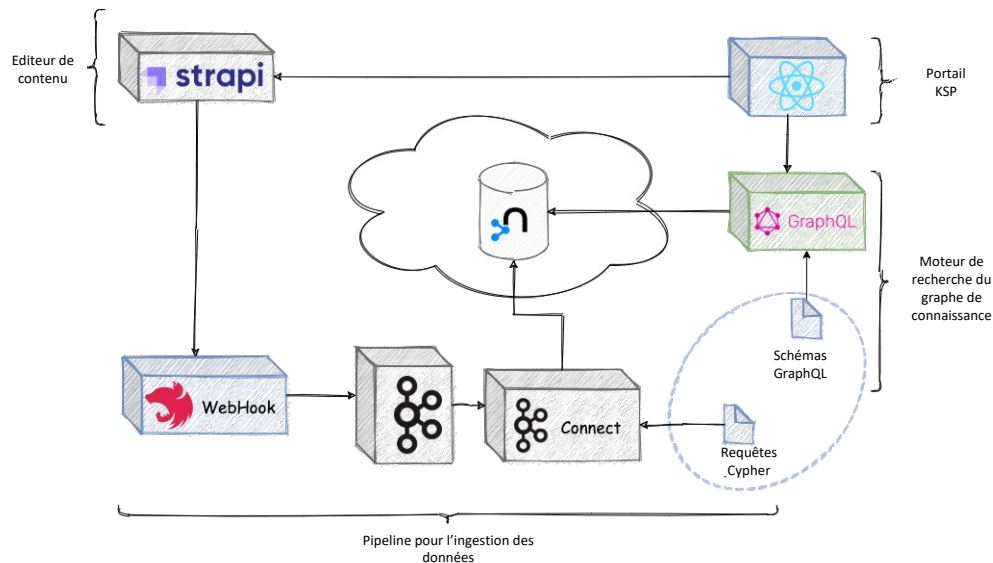


FIG. 3 – Composants logiciels de la plateforme.

Le graphe de connaissance est stocké dans une base graphe Neo4J⁸. Il ne contient que les métadonnées associées aux documents et non les documents eux-mêmes.

Un webservice basé sur la librairie Neo4J GraphQL⁹ sert à exposer le contenu du graphe via une API GraphQL. GraphQL est un langage de requête et un moteur d’exécution qui permet de créer rapidement des APIs flexibles et faciles à utiliser. Notre service GraphQL utilise des fichiers définissant le mapping entre la structure du graphe et des schéma GraphQL pour générer automatiquement une API utilisable par l’interface web.

L’interface de la plateforme, développée avec React¹⁰, utilise donc cette API GraphQL pour requêter la base graphe et afficher les résultats. C’est la partie publique de la plateforme destinée aux membres de la communauté. Elle leur permet d’accéder à des vues prédéfinies du graphe et de rechercher des documents sur la base des concepts de l’ontologie SROnto.

Finalement, l’édition du contenu du graphe peut se faire via le logiciel Strapi¹¹, un headless CMS qui fournit des formulaires de saisie pour les types de concepts principaux du graphe.

8. <https://neo4j.com>

9. https://neo4j.com/developer/graphql/#_the_neo4j_graphql_library

10. <https://reactjs.org>

11. <https://strapi.io>

Une plateforme de gestion des connaissances

Seuls les utilisateurs ayant le droit d'ajouter et de corriger le contenu du graphe ont accès à cet outil. Un pipeline basé sur Kafka permet d'injecter les données publiées via Strapi dans le graphe, et donc dans la plateforme.

6 Evaluation des composants de la plateforme

Pour tester la plateforme KSP, nous proposons d'une part l'évaluation de l'ontologie et d'autre part l'évaluation du graphe de connaissance.

6.1 Evaluation de l'ontologie SROnto

Tout d'abord, nous avons utilisé OOPS! (Poveda-Villalón et al., 2014) pour identifier les erreurs les plus courantes commises lors de la conception d'ontologies. Ensuite, Pour SROnto, nous avons décidé de baser notre évaluation suivant deux approches : la première basée sur les tâches (ou sur les applications) et, la seconde, l'évaluation par des humains. L'ontologie étant construite pour répondre à des questions sur le domaine, SROnto a été évaluée pour sa capacité à répondre à un ensemble initial de questions générées par des experts du domaine. Ceci s'est matérialisé via un ensemble de requêtes SPARQL et la réponse à ces requêtes. De plus, comme l'ontologie sert à indexer les documents à stocker dans la KSP, nous avons vérifié que les documents indexés pouvaient être récupérés via la plateforme si les mots clés composant les requêtes contiennent des étiquettes de concept de SROnto. Deuxièmement, des experts du domaine de l'ESRIC et de l'ESA ont été impliqués dans le processus de validation. Guidé par des ingénieurs de la connaissance, des chercheurs, ingénieurs et des responsables de l'informatique décisionnelle disposant d'une large expérience dans les aspects scientifiques, technologiques et commerciaux des ressources spatiales ont revu et corrigé un par un les entités présentes dans l'ontologie. Quelques itérations ont conduit à une première version publiable de SROnto en 2022.

6.2 Evaluation du graphe de connaissance

Le graphe des connaissances contenant les métadonnées associées aux documents scientifiques et technologiques pertinents du domaine a fait l'objet d'une démonstration lors de la Space Ressources Week 2022. Le graphe de connaissance, visualisable sur une table tangible, a été consulté et critiqué par les 250 participants à cet événement majeur pour la communauté des ressources spatiales. À la suite de ces résultats encourageants, nous poursuivons les évaluations du contenu du graphe notamment par la quantité de doublons, d'information incomplète (e.g., quantité de documents pour lesquels il manque des informations telles que les auteurs ou le type de document) ou d'information contradictoire (i.e., des informations sur le même document qui ne seraient pas les mêmes comme des auteurs différents pour un même article).

7 Conclusion

L'identification et l'utilisation des ressources spatiales deviennent rapidement une réalité, poussées par une évolution de la technologie, l'accélération de l'exploration de l'espace ex-

traatmosphérique, et la rareté de certaines ressources sur Terre. Dans cet article nous avons présenté une plateforme de management des connaissances des ressources spatiales contenu dans les documents pertinents du domaine produits par le monde académique, l'industrie, et les institutions gouvernementales et non-gouvernementales basée sur les technologies du Web Sémantique. Cet outil, accessible via le Web à un grand nombre d'utilisateurs, offre des moyens de cartographier l'état des connaissances sur des thématiques spécifiques du domaine et d'en identifier les acteurs majeurs. Ces travaux s'inscrivant dans la durée, une tâche majeure va consister d'une part à poursuivre le développement de services autour de cette plateforme (e.g., visualisation, système de questions/réponses, partage et création de connaissance) et, d'autre part, à continuer le travail de fédération et d'implication de la communauté dans le développement de la plateforme KSP.

Références

- Arp, R., B. Smith, et A. D. Spear (2015). *Building ontologies with basic formal ontology*. Mit Press.
- Auer, S., M. Stocker, L. Vogt, G. Fraumann, et A. Garatzogianni (2021). ORKG : Facilitating the transfer of research results with the open research knowledge graph. *Research Ideas and Outcomes* 7, e68513.
- Bizer, C., T. Heath, et T. Berners-Lee (2011). Linked data : The story so far. In *Semantic services, interoperability and web applications : emerging concepts*, pp. 205–227. IGI global.
- Burnham, J. F. (2006). Scopus database : a review. *Biomedical digital libraries* 3(1), 1–8.
- Cardoso, S. D., M. Da Silveira, et C. Pruski (2020). Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies. *Knowledge-Based Systems* 194, 105508.
- Da Silveira, M., L. Deladiennee, C. Pruski, et E. Scolan (2022). A knowledge sharing platform for space resources. In *Proceedings of the Space Resources Week*, Luxembourg.
- ESA (2019). ESA space resources strategy. Technical report, European Space Agency.
- Garijo, D. (2017). WIDOCO : a wizard for documenting ontologies. In *International Semantic Web Conference*, pp. 94–102. Springer.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition* 5(2), 199–220.
- Hendricks, G., D. Tkaczyk, J. Lin, et P. Feeney (2020). Crossref : The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies* 1(1), 414–427.
- ISECG (2021). ISECG annual report. Technical report, International Space Exploration Coordination Group.
- Koul, A., S. Ganju, M. Kasam, et J. Parr (2020). Space ML : distributed open-source research with citizen scientists for the advancement of space technology for NASA. *CoRR abs/2012.10610*.
- Liu, Z., M. Shi, K. Janowicz, B. Regalia, S. Delbecque, G. Mai, R. Zhu, et P. Hitzler (2022). LD Connect : A linked data portal for ios press scientometrics. In *European Semantic Web Conference*, pp. 323–337. Springer.

- Lo, K., L. L. Wang, M. Neumann, R. Kinney, et D. S. Weld (2019). S2ORC : The semantic scholar open research corpus. *arXiv preprint arXiv :1911.02782*.
- Meurisse, A., J. Mousel, A. Kapoglou, M. Conti, A. Makaya, A. Cowley, J. Carpenter, M. Link, et B. Hufenbach (2021). Utilization scenarios—outcome of the space resources week 2021. *LPI Contributions 2635*, 5024.
- Poveda-Villalón, M., A. Gómez-Pérez, et M. C. Suárez-Figueroa (2014). OOPS! (Ontology Pitfall Scanner!) : An On-line Tool for Ontology Evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS) 10(2)*, 7–34.
- Pruski, C., N. Guelfi, et C. Reynaud (2011). Adaptive ontology-based web information retrieval : The target framework. *International Journal of Web Portals (IJWP) 3(3)*, 41–58.
- Rovetto, R. (2017). Ontology-based knowledge management for space data. In *68th International Astronautical Congress, Adelaide, Australia*.
- Stephenson, P. (2001). The JORC code. *Applied Earth Science 110(3)*, 121–125.
- Suárez-Figueroa, M. C., A. Gómez-Pérez, et M. Fernández-López (2012). The NeOn methodology for ontology engineering. In *Ontology engineering in a networked world*, pp. 9–34. Springer.
- Vandenbussche, P.-Y., G. A. Ateazing, M. Poveda-Villalón, et B. Vatant (2017). Linked Open Vocabularies (LOV) : a gateway to reusable semantic vocabularies on the web. *Semantic Web 8(3)*, 437–452.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific data 3(1)*, 1–9.
- York, D. G., J. Adelman, J. E. Anderson Jr, S. F. Anderson, J. Annis, N. A. Bahcall, J. Bakken, R. Barkhouser, S. Bastian, E. Berman, et al. (2000). The sloan digital sky survey : Technical summary. *The Astronomical Journal 120(3)*, 1579.

Summary

The ever-increasing interest from academia, industry, and governmental institutions in space resources information has highlighted the difficulty of finding, accessing, integrating, and reusing the related knowledge. All this knowledge is disseminated in scientific publications, patents, news, and reports, which are regularly made available to the community on the Web through many media and in different formats. In collaboration with the European Space Agency, we are currently developing a knowledge management platform in the field of space resources. This tool, based on the use of Semantic Web technologies such as knowledge graphs and ontologies, aims to formalize the knowledge of the domain contained in the documents mentioned above and to make them available to the community. In this article, we describe the concepts and techniques of knowledge extraction and knowledge management implemented for the design and implementation of such a platform.